Выбор функции потери в задаче сегментации спутниковых снимков с использованием сверточной нейронной сети

А.Г.Седов, В.В. Хрящев, Р.В. Ларионов

Ярославский государственный университет им. П.Г. Демидова,

Ярославль, Россия

agsedov@gmail.com, [v.khryashchev@uniyar.ac.ru](mailto:v.khryashchev@uniyar.ac.ru), r.larionov@uniyar.ac.ru

А.А. Островская

Российский университет дружбы народов

(РУДН)

Москва, Россия,

ostrovskaya-aa@rudn.ru

*Аннотация* — Представлены результаты обучения сверточной нейронной сети для сегментации спутниковых снимков. Входные изображения содержат четыре канала: красный, зеленый, синий и ближний инфракрасный. Сверточная нейронная сеть была обучена детектировать здания и сооружения. Для решения задачи сегментации спутниковых снимков использовалась архитектура U-Net. Для обучения использовался суперкомпьютер NVIDIA DGX-1. Был описан процесс аугментации данных, сравниваются результаты обучения с различными функциями потерь. Представлены результаты оценки сети для различных типов жилых районов.

Ключевые слова — спутниковые снимки, сегментация изображений, аугментация данных, функция потерь, архитектура сети U-Net.

# ВВЕДЕНИЕ

Алгоритмы глубокого обучения стали очень популярными в последние годы из-за увеличения мощности вычислительных устройств и усовершенствования технологии создания графических процессоров. Данные дистанционного зондирования предоставляют проблемы для глубокого обучения, потому что анализ спутниковых изображений поднимает уникальные проблемы, которые ставят сложные научные вопросы [1-3].

Одной из наиболее распространенных задач, связанных с анализом спутниковых изображений, является их автоматическая сегментация. В настоящее время для решения этой задачи используются сверточные нейронные сети. Одной из таких архитектур нейронных сетей является архитектура U-Net. В архитектуре U-Net используются пропускаемые соединения для объединения признаков, полученных на разных этапах работы сети. Эта сверточная нейронная сеть показала свою эффективность в сегментации медицинских изображений [4], SAR [5-6] и др.

В этой статье показываются различные этапы обучения нейронных сетей для сегментации спутниковых снимков с целью выделения зданий и сооружений на исходном наборе данных. Результаты такой сегментации могут быть использованы для выделения полигонов, связанных со зданиями.

Эта статья продолжает исследования [7-10], посвященные сегментации спутниковых изображений. В работе [7] было показано, что сверточная нейронная сеть может эффективно использоваться для сегментации спутниковых изображений. Три архитектуры CNN были проанализированы (U-Net, LinkNet и SegNet) для обнаружения классов «водные ресурсы», «лес» и «сельское хозяйство» [8]. Также была проведена оптимизация обхода маски изображения для получения лучших результатов. U-Net показал лучший результат для всех классов: точность классификации составила 81,7% для «воды», 92,3% для «леса» и 96,1% для «сельского хозяйства» [9]. В статье [10] было выполнено обнаружение зданий на спутниковых снимков из базы данных PlanetScope с пространственным разрешением 0,5 м/пиксель.

Некоторые требования к алгоритмам сегментации спутникового изображения [5-10]:

* Размер и тип классифицируемых объектов одного и того же класса могут сильно различаться. Эта проблема может быть решена путем разделения класса по размеру или типу и обучению различных кодировщиков. Разделение по размеру является предпочтительным, потому что это может быть сделано автоматически, без ручного изменения меток набора данных.
* Высокая плотность целевых областей. Например, два здания могут стоять очень близко друг к другу. Алгоритм сегментации должен быть оштрафован за плохое разделение объектов во время обучения для улучшения качества выходных масок. Это достигается путем тщательного выбора функции потерь.
* Небольшой размер базы данных. Успешность обучения алгоритма машинного обучения во многом зависит от качества и количества обучающих данных. В этой статье целью было достижение наилучших результатов с оригинальными спутниковыми изображениями, поэтому предпринимались попытки преодолеть ограничения набора данных. Одним из способов максимально использовать минимальный набор данных является аугментация данных. Другой способ - использовать большие наборы данных для предварительного обучения сети, а затем подстройка сверточных нейронных сетей на целевые данные.
* Результаты обучения должны быть инвариантны к поворотам, в то время как некоторые объекты могут присутствовать в наборе данных обучения только один раз. Эта проблема также может быть решена путем аугментации данных.
* Пространственное разрешение спутниковых изображений также может варьироваться.

978-1-7281-3238-9/19/$31.00 ©2019 IEEE

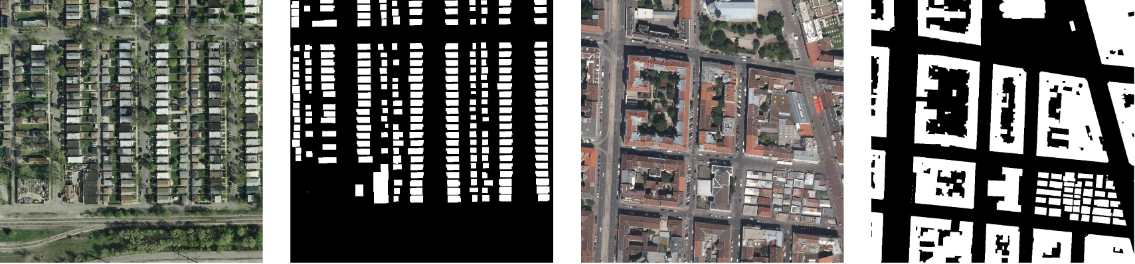


Рис. 1. Примеры спутниковых снимков и масок базы данных Inria

Эта статья состоит из шести частей. В первой части читателя знакомят с поставленной задачей. Во второй части описываются используемые наборы данных. В третьей части приводится архитектура сети, а также описываются протестированные функции потерь. В четвертой части представлены результаты численных экспериментов. Последняя часть предназначена для выводов и будущих направлений работы.

# НАБОР ОБУЧАЮЩИХ ДАННЫХ

База данных Inria [3] содержит цветные спутниковые снимков, покрывающие территорию суммарной площадью 810 км². Обучающая и тестовая выборка составляет 180 изображений размером 1000x1000 пикселей при пространственном разрешении 0,3 м/пиксель. Все снимки разделены на 2 класса: «здания» и «не здания». Изображения из базы данных Inria охватывают разные городские поселения, начиная от густонаселенных районов и заканчивая альпийскими городами: Сан-Франциско (США), Чикаго (США), Вена (Австрия), Инсбук (Австрия), Беллингэм (США) и Тироль (Австрия). Примеры изображений из базы данных Inria показаны на рис. 1.

Набор данных, на которые проходило обучение модели, состоит из изображений 16 регионов, от сельских до городских районов. Каждое изображение имеет свою двоичную маску, размеченную экспертами. Набор данных охватывает 25 квадратных километров. Эти изображения состоят из четырех каналов: красного, зеленого, синего и ближнего инфракрасного (NIR) и немного отличаются друг от друга по пространственному разрешению.

U-Net в качестве входных данных принимает изображения размером 256 x 256 пикселей. Для создания обучающего и валидационного набора данных, каждое изображение из базы данных было порезано на две непересекающиеся полосы, а затем каждая полоса была разделена на перекрывающиеся фрагменты 256 x 256 пикселей с шагом в 128 пикселей. Такое разделение применялось к каналам RGB и NIR по отдельности.

Потом, для увеличения обучающего набора данных, применялось следующее:

1. Повороты на 90, 180, 270 градусов и зеркальные отражения. В результате, после этих преобразований обучающий набор данных увеличился в 8 раз.

2. Хроматические искажения. Изображения были переведены из цветового пространства RGB в цветовое пространство HSV и к новым цветовым координатам HSV были добавлены случайные значения. Для канала NIR вместо хроматических искажений к нормализованным в интервале [0, 1] значениям были добавлены случайные величины из интервала [-0,06, +0,06].

3. Случайные сдвиги, перемасштабирование и повороты малых степеней.

После каждого этапа аугментации проводилась оценка эффективности сегментации на валидационном наборе данных. Эксперименты осуществлялись с различными значениями параметров хроматических искажений, случайных сдвигов и вращений для достижения наилучшего результата. Для оценки сходства результатов работы алгоритма с размеченными масками использовался коэффициент подобия Серенсена

Мы использовали метрику в кости, чтобы оценить сходство между предсказанными изображениями и сделанными экспертом масками.

# АРХИТЕКТУРА СЕТИ

Нами была использована модификация сети U-Net для задачи сегментации данных. Классическая сеть U-Net состоит из 2 частей: кодировщик и декодировщик (рис. 2).

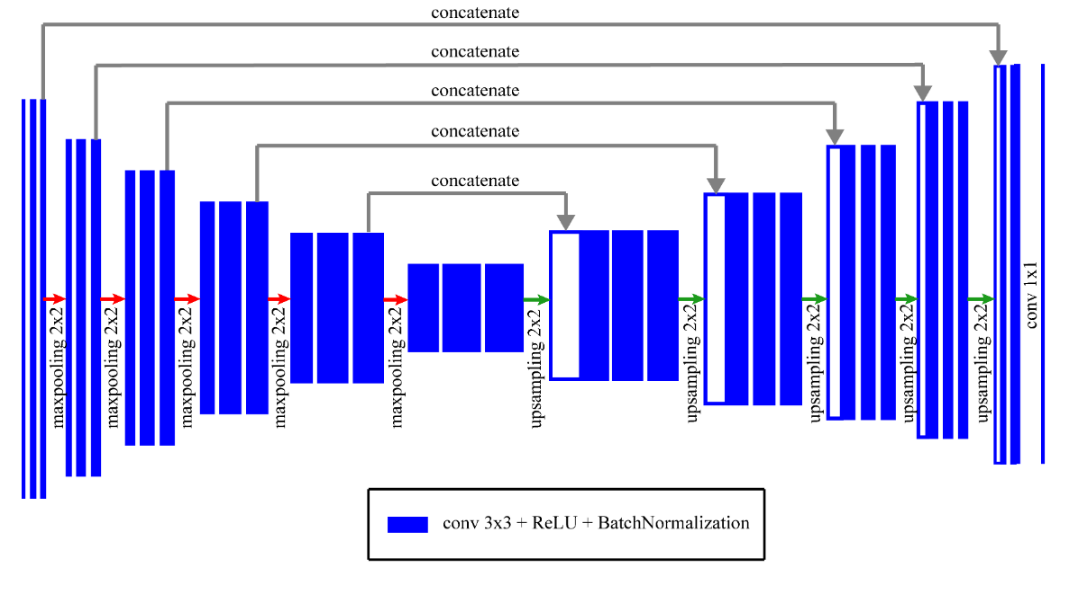


Рис. 2. Архитектура сети U-Net

Кодировщик представляет собой сверточную нейронную сеть, состоящую из четырех блоков. Каждый такой блок состоит из двух сверточных слоев с фильтрами 3 × 3, с функцией активации ReLU и пакетной нормализацией, применяемой к каждому из них, а также слоя понижения дискретизации с окном 2 × 2. Декодировщик имеет такое же количество блоков, что и кодировщик. Каждый блок декодировщика состоит из слоя повышения дискретизации с окном 2 × 2, слияния с соответствующим набором признаков из кодировщика, двух сверточных слоев с фильтрами 3 × 3 и функцией активации ReLU, примененной к каждому из них. Последний слой использует сигмоидную функцию активации признаков для классификации на уровне пикселей.

Оригинальная архитектура сети U-Net была преобразована: в нашем случае имели место два отдельных кодировщика для компонентов RGB и NIR. Выходы декодировщика объединялись перед присоединением к центральному и верхнему уровням сети. Для обучения модели использовался оптимизатор Adam [12].

Сеть была обучена с использованием четырех функций потерь:

## Взвешенная бинарная кросс-энтропия со взвешенными коэффициентами

Эта функция потерь вычисляется по следующей формуле:

*.*

Здесь WBCE и WDICE обозначают взвешенное значение BCE and коэффициент Серенсена:

.

Функция в случае, если пиксель был классифицирован верно и в противном случае, а если указанный пиксель относился к указанному классу в соответствии с данными наземного контроля данных.

Взвешенная функция , появлявшаяся в WBCE и WDICE – единственное, что отличает их от оригинальных функций DICE и BCE. Эта функция устанавливает приоритет пикселям рядом с границами объекта над другими пикселями.

## Бинарная кросс-энтропия с оригинальными значениями весов и коэффициентом Серенсена

Формула взвешенной бинарной кросс-энтропии была преобразована согласно статье [4]. Чтобы вычислить это сначала, вычисляются значения весов:

,

- карта весов для балансировки частот классов,

и обозначают расстояния до ближайших объектов.

и σ – глобальные параметры.

Затем, для каждого пикселя вычисляется функция softmax:

,

где обозначает активацию в канале объекта в позиции . Функция энергии оценивается с использованием бинарной кросс-энтропии.

И наконец функция потерь рассчитывается по следующей формуле:

.

## Бинарная кросс-энтропия с оригинальными значениями весов

То же самое, что описано выше, за исключением выражения в скобках с коэффициентом Серенсена:

## Функция потерь Lovász-Softmax.

Была также реализована модифицированная функция потерь из статьи [11]:

,

где LH обозначает фунцию потерь Lovász Hinge [15].

И хотя значение коэффициента Серенсена сильно не улучшилось, сеть U-Net, обученная с помощь функции потерь Lovász-Softmax, показала наилучшие результаты по детектированию зданий (рис. 3).

# РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

Пример маски, полученной после проведения аугментации обучающих данных, приведен на рис. 4. Во время предварительного обучения значение коэффициента Серенсена равного 0,84 было достигнуто на наборе данных Inria для RGB-кодировщика. Улучшение значения коэффициента Серенсена после выполнения различных этапов аугментации данных можно увидеть на рис. 5. Улучшение значения данной метрики варьируется для спутниковых снимков различных регионов.

Результаты применения различных функций потерь для сети U-Net также были проанализированы. В результате было установлено, что значение коэффициента Серенсена варьируется в диапазоне от 0.73 до 0.75.

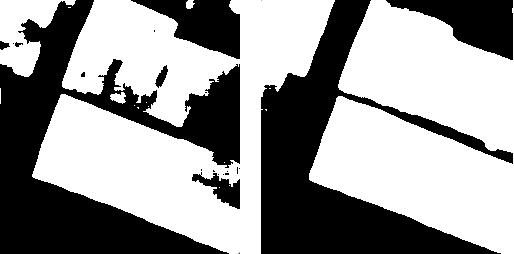


Рис. 3. Предсказания для модели, обученной с помощью взвешенной бинарной кросс-энтропии (слева) и с помощью функции потерь Lovász-Softmax (справа)



Рис. 4.Размеченные маска и результат алгоритма после аугментации обучающих данных

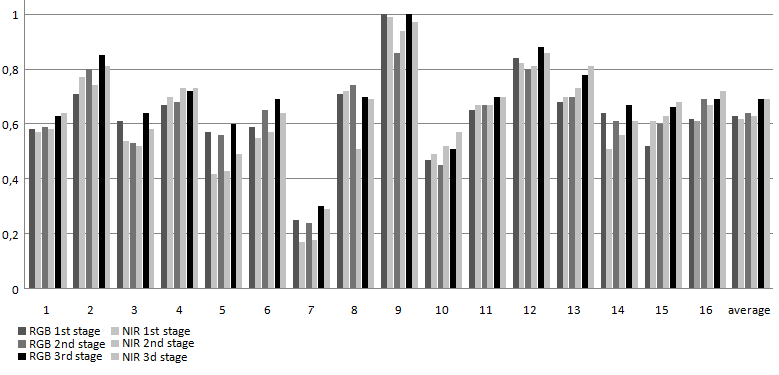


Рис. 5. Значение коэффицента Серенсена для 16 различных регионов для RGB и NIR после выполнения 3 этапов аугментации данных.

# ЗАКЛЮЧЕНИЕ

В данной статье описан процесс обучения нейронной сети, предназначенной для сегментации спутниковых изображений. Коэффициент Серенсена имеет потенциал для дальнейшего улучшения после проведения дополнительных экспериментов с предварительным обучением сети на другом наборе данных и ее настройке.

Другими направлениями исследования могут быть:

* Детектирование зданий разных размеров разными кодировщиками
* Предварительное обучение сети на других базах данных, например Imagenet.
* Экспериментыв с различными методами объединения каналов NIR и RGB.

Разработанный алгоритм может найти свое применение при решении задачи по оценке уровня урбанизации различных регионов.

##### Благодарности

Статья была подготовлена при финансовой поддержке Министерства образования Российской Федерации в рамках научного проекта № 14.575.21.0167 (идентификатор RFMEFI57517X0167).

##### Список литературы

1. Zhang, Liangpei, Lefei Zhang, and Bo Du. “Deep learning for remote sensing data: A technical tutorial on the state of the art.” IEEE Geoscience and Remote Sensing Magazine 4.2 (2016): 22-40.
2. Y. Chen, H. Jiang, C. Li, X. Jia and P. Ghamisi, “Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks,” in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232-6251, Oct. 2016.
3. X. X. Zhu et al., “Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources,” in IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8-36, Dec. 2017.
4. O.Ronneberger, P. Fischer “U-Net: Convolutional Networks for Biomedical Image Segmentation”, arXiv:1505.04597. 2015
5. Zhang, Z., Liu, Q., & Wang, Y. (2018). “Road extraction by deep residual u-net”. IEEE Geoscience and Remote Sensing Letters, 15(5), pp. 749-753.
6. Iglovikov, V., Mushinskiy, S., & Osin, V. (2017). “Satellite imagery feature detection using deep convolutional neural network: A kaggle competition”. arXiv preprint arXiv:1706.06169.
7. V. V. Khryashchev, V. A. Pavlov, A. Priorov and A. A. Ostrovskaya, "Deep Learning for Region Detection in High-Resolution Aerial Images," 2018 IEEE East-West Design & Test Symposium (EWDTS), Kazan, 2018, pp. 1-5.
8. V.Khryashchev, L.Ivanovsky, V.Pavlov, A.Ostrovskaya, A.Rubtsov Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation, Proceedings of the FRUCT’23, Bologna, Italy, 13–16 November 2018. pp.172–179.
9. V. V. Khryashchev, A. A. Ostrovskaya, V. A. Pavlov and A. S. Semenov, "Optimization Of Convolutional Neural Network For Object Recognition On Satellite Images," Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Minsk, 2018, pp. 1-5.
10. L. Ivanovsky, V. Khryashchev, V. Pavlov and A. Ostrovskaya, "Building Detection on Aerial Images Using U-NET Neural Networks," 2019 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 2019, pp. 116-122.
11. E.Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez, “Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark”, IEEE International Geoscience and Remote Sensing Symposium (IGARSS). 2017.
12. D. P. Kingma, J. Ba, “Adam: A Method for Stochastic Optimization”, Web: https://arxiv.org/abs/1412.6980.
13. S.Golovanov, R. Kurbanov, A. Artamonov, A. Davydow, S. Nikolenko Building Detection From Satellite Imagery Using a Composite Loss Function, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018
14. M.Berman, A. Triki., M. Blaschko. “The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks” // arXiv:1705.08790, 2017.
15. R. Hamaguchi, S. Hikosaka,Building Detection From Satellite Imagery Using Ensemble of Size-Specific Detectors; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018, pp. 187-191